The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 146. The invention includes a fragment of a polypeptide comprising SEQ ID NO: 146. The invention includes a diagnostic kit comprising a polypeptide comprising SEQ ID NO: 146, or a fragment thereof. The invention includes a diagnostic kit comprising a polynucleotide sequence encoding SEQ ID NO: 146, or a fragment thereof. The invention includes an immunogenic composition comprising a polypeptide comprising SEQ ID NO: 146, or a fragment thereof. The invention includes an antibody which recognizes a polypeptide comprising SEQ ID NO: 146, or a fragment thereof.

SEQ ID NO: 146 also contains an open reading frame comprising SEQ ID NO: 147. The invention includes a polypeptide comprising SEQ ID NO: 147. SEQ ID NO: 147 is set forth below.

SEQ ID NO: 147

MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLFLP
FYSNVTGFHTINHTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQSVIIINNSTN
VVIRACNFELCDNPFFAVSKPMGTQTHTMIFDNAFNCTFEYISDAFSLDVSEKSGNFKHL
REFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINITNFRAILTAFSPAQDI
WGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNPLAELKCSVKSFEIDKGIYQTS
NFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFST
FKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVL
AWNTRNIDATSTGNYNYKYRLFRHGKLRPFERDISNVPFSPDGKPCTPPALNCYWPLND
YGFYTTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTP
SSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCAFGGVSVITPGTNASSEVAVLYQDV
NCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEHVDTSYECDIPIGAGICASYH
TVSLLRSTSQKSIVAYTMSLGADSSIAYSNNTIAIPTNFSISITTEVMPVSMAKTSVDCNMY
ICGDSTECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQMYKTPTLKYFGGF
NFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMKQYGECLGDINARDLICAQKFNGLTVL
PPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYEN
QKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDI
LSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRV
DFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHEGKAYFPREGVFVFNG
TSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTS
PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVWLGFI
AGLIAIVMVTILLCCMTSCCSCLKGACSCGSCCKFDEDDSEPVLKGVKLHYT

The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 147. The invention includes a fragment of a polypeptide comprising SEQ ID NO: 147. The invention includes a diagnostic kit comprising a polypeptide comprising SEQ ID NO: 147, or a fragment thereof. The invention includes a diagnostic kit comprising a polynucleotide sequence encoding SEQ

ID NO: 147, or a fragment thereof. The invention includes an immunogenic composition comprising a polypeptide comprising SEQ ID NO: 147, or a fragment thereof. The invention includes an antibody which recognizes a polypeptide comprising SEQ ID NO: 147, or a fragment thereof. SEQ ID NO: 147 demonstrates functional homology to a coronavirus spike protein.

Predicted transmembrane regions of SEQ ID NO: 147 are identified below.

Predicted Transmembrane helices of SEQ ID NO: 147

```
The sequence positions in brackets denominate the core region.
Only scores above 500 are considered significant.

Inside to outside helices :  18 found
      from         to     score center
   1 (   1)  16 (  16)     959      9
 233 ( 237) 257 ( 252)     905    244
 345 ( 347) 364 ( 361)     490    354
 345 ( 354) 369 ( 369)     420    362
 497 ( 497) 513 ( 513)     239    506
 573 ( 573) 588 ( 588)     811    580
 645 ( 648) 666 ( 663)     302    656
 690 ( 696) 714 ( 711)     428    704
 857 ( 860) 882 ( 874)    1508    867
1031 (1031)1046 (1046)     446   1039
1199 (1203)1219 (1217)    2667   1210

Outside to inside helices :  13 found
      from         to     score center
   1 (   1)  17 (  17)     684     10
 222 ( 222) 240 ( 237)     238    229
 244 ( 247) 264 ( 264)     613    254
 349 ( 355) 369 ( 369)     314    362
 496 ( 496) 511 ( 511)     488    503
 573 ( 573) 591 ( 591)     712    581
 650 ( 652) 666 ( 666)     474    659
 674 ( 679) 702 ( 696)     190    686
 691 ( 696) 713 ( 711)     210    704
 866 ( 868) 886 ( 886)    1172    876
1198 (1201)1215 (1215)    3221   1208
```

SEQ ID NO: 147, the spike protein, is a surface exposed polypeptide. Recombinant expression of a protein can be hindered by hyrophobic transmembrane regions. Accordingly, the invention includes a polypeptide comprising SEQ ID NO: 147 wherein one or more of the hydrophobic regions identified above is removed. The invention further includes a polynucleotide encoding such a polypeptide. The invention includes recombinantly expressing the protein in a host cell.

Further characterization of SEQ ID NO: 147 is set forth below.

PSORT --- Prediction of Protein Localization Sites

```
version 6.4(WWW)
MYSEQ          1255 Residues
Species classification: 4

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
     count: 2
     Position of the most N-terminal TMS: 496 at i=2
MTOP: membrane topology (Hartmann et al.)
     I(middle): 503   Charge diffirence(C-N):  1.0
McG: Examining signal sequence (McGeoch)
     Length of UR:   13
     Peak Value of UR:   3.28
     Net Charge of CR: 0
     Discriminant Score:       8.66
GvH: Examining signal sequence (von Heijne)
     Signal Score (-3.5): 5.94
     Possible cleavage site: 13
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition of Predicted Mature Form:
   calculated from 14
ALOM new cnt: 1 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 0B
ALOM: finding transmembrane regions (Klein et al.)
     count: 1  value: -12.26  threshold: -2.0
     INTEGRAL    Likelihood =-12.26   Transmembrane 1202 -1218 (1194 -
1228)
     PERIPHERAL  Likelihood =  0.16
     modified ALOM score:   2.55
>>> Seems to be a Type Ia membrane protein
     The cytoplasmic tail is from 1219 to 1255 (37 Residues)
Rule: vesicular pathway
Rule: vesicular pathway
Rule: vesicular pathway
(14) or uncleavable?
Gavel: Examining the boundary of mitochondrial targeting seq.
     motif at: 14
     Uncleavable? Ipos set to: 24
Discrimination of mitochondrial target seq.:
     positive ( 2.18)
Rule: vesicular pathway
Rule: vesicular pathway
Rule: vesicular pathway

*** Reasoning Step: 2

KDEL   Count: 0
Checking apolar signal for intramitochondrial sorting
  (Gavel position 24)  from: 1  to: 10  Score:  8.0
SKL motif (signal for peroxisomal protein):
     pos: 964(1255), count: 1   SRL
     SKL score (peroxisome):  0.1
Amino Acid Composition Tendency for Peroxisome:   1.37
     AAC not from the N-term., score modified
```

```
Peroxisomal proteins?   Status: notclr
     AAC score (peroxisome): 0.079
Amino Acid Composition tendency for lysosomal proteins
     score:  0.39  Status: notclr
GY motif in the tail of typeIa? (lysosomal)
Checking the amount of Basic Residues (nucleus)
Checking the 4 residue pattern for Nuclear Targeting
Checking the 7 residue pattern for Nuclear Targeting
Checking the Robbins & Dingwall consensus (nucleus)
Checking the RNA binding motif (nucleus or cytoplasm)
Nuclear Signal   Status: negative ( 0.00)
Type Ia is favored for plasma memb. proteins
Checking the NPXY motif..
Checking the YXRF motif..
Checking N myristoylation..

----- Final Results -----

plasma membrane --- Certainty= 0.460(Affirmative) < succ>
microbody (peroxisome) --- Certainty= 0.171(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty= 0.100(Affirmative) <
succ>
endoplasmic reticulum (lumen) --- Certainty= 0.100(Affirmative) < succ>
```

SEQ ID NO: 147 appears to have a N-terminus signaling region, followed by a surface exposed region, followed by a transmembrane region followed by a C-terminus cytoplasmic domain region. Accordingly, the invention includes an immunogenic, surface exposed fragment of SEQ ID NO: 147. Preferably, said fragment comprises an amino acid sequence which does not include the last 50 amino acids of the C-terminus of SEQ ID NO: 147. Preferably, said fragment comprises an amino acid sequence which does not include the last 70 amino acids of the C-terminus of SEQ ID NO: 147. Preferably, said fragment does not include a transdomain region of SEQ ID NO: 147. Preferably, said fragment does not include a C-terminus cytoplasmic domain of SEQ ID NO: 147. Preferably, said fragment does not include a N-terminus signal sequence. Preferably, said fragment does not include amino acids 1 – 10 of the N-terminus of SEQ ID NO:147. Preferably, said fragment does not include amino acids 1 – 14 of the N-terminus of SEQ ID NO: 147.

The spike protein of coronaviruses may be cleaved into two separate chains into S1 and S2. The chains may remain associated together to form and form a dimer or a trimer. Accordingly, the invention includes a polypeptide comprising SEQ ID NO: 147 wherein said polypeptide has been cleaved into S1 and S2 domains. The invention further includes a polypeptide comprising SEQ ID NO: 147 wherein amino acids 1 – 10,

preferably amino acids 1 – 14 of the N-terminus are removed and further wherein SEQ ID NO: 147 is cleaved into S1 and S2 domains. Preferably the polypeptide is in the form of a trimer.

Predicted N-glycosylation sites of SEQ ID NO: 147 are identified below:

Prediction of N-glycosylation sites of SEQ ID NO: 147

SeqName	Position	Potential	Jury agreement	NGlyc result
SEQID 147	29 NYTQ	0.7751	(9/9)	+++
SEQID 147	65 NVTG	0.8090	(9/9)	+++
SEQID 147	109 NKSQ	0.6081	(7/9)	+
SEQID 147	119 NSTN	0.7039	(9/9)	++
SEQID 147	158 NCTF	0.5808	(7/9)	+
SEQID 147	227 NITN	0.7518	(9/9)	+++
SEQID 147	269 NGTI	0.6910	(9/9)	++
SEQID 147	318 NITN	0.6414	(9/9)	++
SEQID 147	330 NATK	0.6063	(8/9)	+
SEQID 147	357 NSTF	0.5746	(8/9)	+
SEQID 147	589 NASS	0.5778	(6/9)	+
SEQID 147	602 NCTD	0.6882	(9/9)	++
SEQID 147	699 NFSI	0.5357	(7/9)	+
SEQID 147	783 NFSQ	0.6348	(9/9)	++
SEQID 147	1080 NGTS	0.5806	(7/9)	+
SEQID 147	1116 NNTV	0.5106	(5/9)	+
SEQID 147	1176 NESL	0.6796	(9/9)	++

Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 147 wherein said fragment comprises one or more of the glycosylation sites identified above. The invention further includes a polynucleotide encoding one or more of the fragments identified above. This glycosylation site can be covalently attached to a saccharide. Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 147 wherein said fragment comprises one or more of the glycosylation sites identified above and wherein said polypeptide is glycosylated at one or more of the sites identified above.

Predicted O-glycosylation sites are identified below:

Prediction of O-glycosylation sites

Name	Residue	No.	Potential	Threshold	Assignment
SEQID 147	Thr	698	0.8922	0.7696	T
SEQID 147	Thr	706	0.9598	0.7870	T
SEQID 147	Thr	922	0.9141	0.7338	T
SEQID 147	Ser	36	0.8906	0.7264	S
SEQID 147	Ser	703	0.8412	0.7676	S

The invention includes a polypeptide comprising a fragment of SEQ ID NO: 147 wherein said fragment comprises one or more of the o-glycosylation sites identified above. The invention further includes a polynucleotide encoding one or more of the fragments identified above. The invention further includes a polypeptide comprising a fragment of SEQ ID NO: 147 wherein said fragment comprises one or more of the 0-glycosylation sites identified above and further wherein the polypeptide is covalently bonded to a saccharide at one or more of the included glycosylation sites.

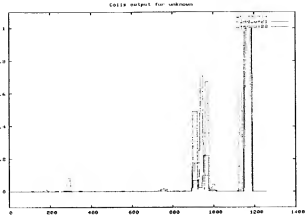
The invention further includes a polypeptide comprising a fragment of SEQ ID NO: 147 wherein said fragment comprises one or more of the N-glycosylation sites identified above and further wherein said fragment comprises one or more of the O-glycosylation sites identified above.

The invention includes a polypeptide comprising a fragment of SEQ ID NO: 147 wherein said fragment does not include one or more of the glycosylation sites identified above. The invention also includes a polynucleotide encoding such a polypeptide.

Predicted phosphorylation sites of SEQ ID NO: 147 are Ser-346, Tyr-195, and Tyr-723. Accordingly, the invention comprises a polypeptide comprising a fragment of SEQ ID NO: 147 wherein said fragment comprises at least ten amino acid residues and wherein said fragment comprises one or more of the amino acids selected from the group consisting of Ser-346, Tyr-195, and Tyr-723. In one embodiment, one or more of the amino acids selected from the group consisting of Ser-346, Tyr-195, and Tyr-723 are phosphorylated.

Predicted coiled coils of SEQ ID NO: 147 are identified below:

Coiled coil Prediction:



Accordingly, the invention comprises a polypeptide sequence comprising a fragment of SEQ ID NO: 147 wherein said fragment includes a coiled region of SEQ ID NO: 147. The invention comprises a polypeptide sequence comprising a fragment of SEQ ID NO: 147, wherein said fragment does not include a coiled region of SEQ ID NO: 147.

The ORF1a and ORF1b sequences of coronaviruses are typically translated as a single ORF1ab polyprotein. Slippage of the ribosome during translation generates an a-1 frameshift. One region of such slippage is illustrated below:

```
   gggttttacacttagaaacacagtctgtaccgtctgcggaatgtggaaaggttatggctgtagttgtga
+1   G  F  T  L  R  N  T  V  C  T  V  C  G  M  W  K  G  Y  G  C  S  C  D
+3  G  F  Y  T  -  K  H  S  L  Y  R  L  R  N  V  E  R  L  W  L  -  L  -
   ccaactccgcgaaccctgatgcagtctgcggatgcatcaacgttttttaaacgggtttgcggtgtaagt
+1   Q  L  R  E  P  L  M  Q  S  A  D  A  S  T  F  L  N  G  F  A  V  -  V
+3  P  T  P  R  T  L  D  A  V  C  G  C  I  N  V  F  K  R  V  C  G  V  S
   gcagcccgtcttacaccgtgcggcacaggcactagtactg
+1   Q  P  V  L  H  R  A  A  Q  A  L  V  L
+3  A  A  R  L  T  P  C  G  T  G  T  S  T
```

which would generate the following translational slippage:

```
ccaactccgcgaaccctgatgcagtctgcggatgcatcaacgttttttaaacgggtttgcggtgtaagt
     Q  L  R  E  P  L  M  Q  S  A  D  A  S  T  F  L  N  R  V  C  G  V  S
```

Accordingly, the invention includes a polypeptide comprising SEQ ID NO: 148. SEQ ID NO: 148 is set forth below.

SEQ ID NO: 148

MESLVLGVNEKTHVQLSLPVLQVRDVLVRGFGDSVEEALSEAREHLKNGTCGLVELEKGVLPQLEQPYV
FIKRSDALSTNHGHKVVELVAEMDGIQYGRSGITLGVLVPHVGETPIAYRNVLLRKNGNKGAGGHSYGI
DLKSYDLGDELGTDPIEDYEQNWNTKHGSGALRELTRELNGGAVTRYVDNNFCGPDGYPLDCIKDFLAR